Soumya Mukherjee

CH24M571

ID5001W

DATE 11/4/25

PLACE Kolkata

ML

UoP Assignment - 2

(1) we have 2 classes $y_i = 1$ & $y_i = -1$

Given 4 points in 2D feature space

$x_1 = (2, 2)$, $y_1 = +1$

$x_2 = (4, 4)$, $y_2 = +1$

$x_3 = (1, -1)$, $y_3 = -1$

$x_4 = (-3, -3)$, $y_4 = -1$

Given normal vector to hyperplane at 0 $\Rightarrow$ $w_0 = (0, 0)$.

Bias at 0 $\Rightarrow$ $b_0 = 0$.

update rule is

$$w^{t+1} = w^t + y_i x_i$$

$$b^{t+1} = b^t + y_i$$

when prediction $\neq y_i$, we need to update weight 'w' & bias 'b'.

DATE

PLACE

we know, Perceptron update rule is

If $w^T x_i + b \geq 0$, predict 1

else if $w^T x_i + b < 0$, predict -1

Epoch - 1

(a) $x_1 = (2, 2)$.

$w.x + b = (0,0).(2,2) + 0$

$= 0 + 0 \rightarrow$ Predict: +1

(correct Prediction)

(b) $x_2 = (4,4)$

$w.x + b = (0,0).(4,4) + 0$

$= 0 + 0 = 0 \rightarrow$ Predict: +1

correct Prediction.

(c) $x_3 = (1, -1)$

$wx + b = (0,0).(1,-1) + 0$

$= 0 \rightarrow$ Predict: +1

Given -1 $\rightarrow$ wrong prediction

(d) $x_4 = (-3, -3)$

$wx + b = (0,0).(-3,-3) + 0$

$= 0 \rightarrow$

Since, wrong prediction, we would want to update weights & bias.

- $w = (0,0) + (1,-1).(-1) = (-1,1)$.
  $b = 0 + (-1) = -1$.

(d) $x_4 = (-3,-3)$.

$wx + b = (-1,1).(-3,-3) + (-1)$
$= 3 + (-3) - 1$
$= -1$ Prediction $= -1$ since < 0.
∴ Correct Prediction

Epoch-2

(a) $x_1 = (2,2), y = +1$ $w = (-1,1)$ $b = -1$

$wx + b = (-1,1).(2,2) + -1$
$= -1$
Wrong Prediction

Performing update operation..

$w = (-1,1) + (+1)(2,2)$
$= (1,3)$
$b = -1 + (+1) = 0$.

(b) $x_2 = (4,4)$, $y = +1$

$wx + b = (1,3).(4,4) + 0$
$= 16 > 0 \Rightarrow$ Predict: 1
This, Prediction is correct

(c) $x_3 = (1,-1)$, $y = -1$

$\Rightarrow wx + b = (1,3).(1,-1) + 0$
$= 1 - 3 + 0 = -2 < 0$.
Predict -1, Prediction: Correct

(d) $x_4 = (-3,-3)$, $y = -1$

$wx + b = (1,3).(-3,-3) + 0$
$= -3 - 9 = -12 < 0$,
Predict: correct (-1)

Thus after 2 updates,
$w = (1,3)$ & $b = 0$

Ans (a)

DATE
PLACE

(b) The data is linearly seperable if

$y(w^T x + b) > 0$ for all points

ie the predicted signs matches the actual class label y for all data points.

using $w = (1,3)$ & $b = 0$, if we are able to correctly predict all the 4 cases, we can claim it to be linearly seperable.

Point	Given label	$wx+b$	Prediction	Correct
(2,2)	+1	$1\times2 + 3\times2 = 8$ >0	+1	Yes
(4,4)	+1	$4\times1 + 3\times4 = 16$ >0	+1	Yes
(1,-1)	-1	$1\times1 + 3\times(-3) = -8$ <0	-1	Yes
(-3,-3)	-1	$1\times-3 + 3\times(-3)$ $= -12$ <0	-1	Yes

DATE
PLACE

Since, all predictions match the labels, the data is linearly seperable by the hyperplane $w^T x_i + b = 0$.

(c) we need to calculate γ-margin for each points.

$$\gamma_i = \frac{y_i(w^T \cdot x_i + b)}{\|w\|}$$

with given $w = (1,3) \Rightarrow \|w\| = \sqrt{1+9} = \sqrt{10}$.

$b = 0$

(a) $x = (2,2)$, $y = +1$

$\gamma_i = (1 \times (1,3)^T \cdot (2,2) + 0)/\sqrt{10}$

$= 8/\sqrt{10}$

(b) $x = (4,4)$, $y = +1$

$\gamma_i = (+1 \times (1,3)^T \cdot (4,4) + 0)/\sqrt{10}$

$= 16/\sqrt{10}$

(c) $x = (1,-1)$, $y = -1$

$\gamma_i = (-1 \times (1,3)^T (1,-1) + 0)/\sqrt{10}$

$= +2/\sqrt{10}$

DATE

PLACE

(a) $x_4 = (-3, -3)$, $y = -1$

$\gamma_i = -1 \times ((1,3)^T \cdot (-3,-3) + 0) / \sqrt{10}$

$= 12/\sqrt{10}$

Thus, among the 4 points given,

(c) $x = (1, -1)$ has the lowest margin.

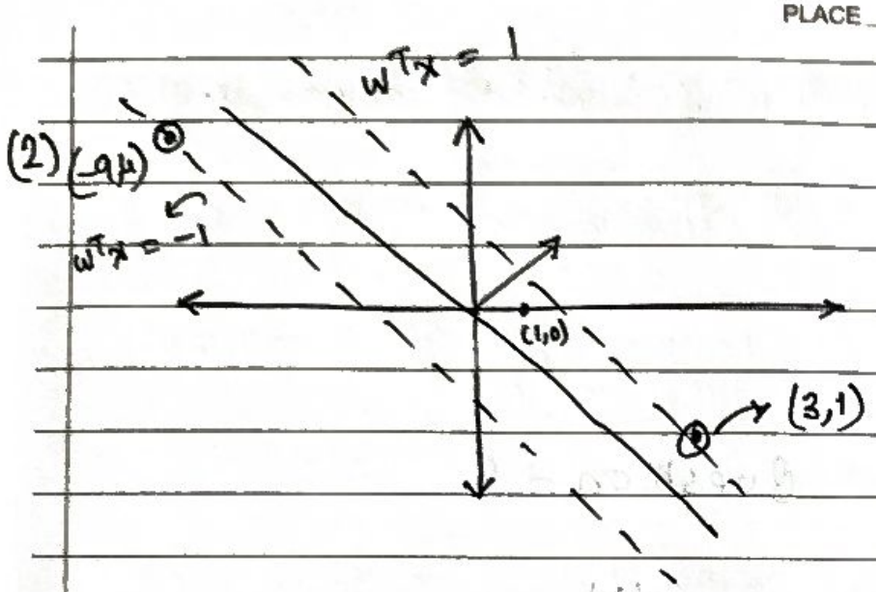
(Ans) ———————————

DATE

PLACE

Question - 2

SVM



(1,0) is a test point.

Given hard-margin SVM i.e perfectly seperates the data.

(a) Supporting hyperplanes are

$w^T x = 1$ for class +1 (green) & $w^T x = -1$ for class -1 (red)

The decision boundary is $w^T x = 0$. which is exactly half way between the two supporting hyperplanes.

(b) width of seperation between 2 hyperplanes is given by,

$$\frac{2}{\|w\|}$$

Given, $(-9,4)$ is on line $w^T x = -1$
& $(3,1)$ is on line $w^T x = 1$.

If $w = [w_1 \; w_2]$

then,

$$[w_1 \; w_2]^T [-9,4] = -1$$

& $$[w_1 \; w_2]^T [3,1] = 1.$$

ie, $-9w_1 + 4w_2 = -1 \quad \rightarrow (1)$

& $3w_1 + w_2 = 1 \quad \rightarrow (2)$

Multiplying (2) by 3 & summing to 1.

$$9w_1 + 3w_2 = 3$$
$$+ \quad 7w_2 = 2$$
$$w_2 = 2/7. \; - 6/21.$$

Thus, $3w_1 = 1 - 2/7 = \frac{5}{7}$

$$w_1 = \frac{5}{21}$$

Thus, $w = \left[\frac{5}{21} \quad \frac{6}{21}\right]$

Calculating L2 norm $\|w\|_2$

$$\|w\| = \sqrt{\left(\frac{5}{21}\right)^2 + \left(\frac{6}{21}\right)^2} = \frac{1}{21}\sqrt{25+36}$$

$$= \frac{1}{21}\sqrt{61}.$$

Thus, width of seperation

$$= \frac{2}{\|w\|} = \frac{42}{\sqrt{61}} \quad \text{(Ans)}$$

→ Equation of decision boundary.

$$w_1 x_1 + w_2 x_2 = 0.$$

or $\frac{5}{21} x_1 + \frac{6}{21} x_2 = 0$

or, $5x_1 + 6x_2 = 0$ (Ans)

K-means

(3)

Initial points

Initial clusters

A1 = (2, 10) → C1 = A1, A4, A7.

A2 = (2,5)

A3 = (8,4)

A4 = (5,8) → C2

A5 = (7,5)

A6 = (6,4)

A7 = (1,2) → C3.

A8 = (4,9).

Calculating distance to each Point

Pt	Dist to A1	Dist to A4	Dist to A7	Assign
A1	0	$\sqrt{13}=3.61$	$\sqrt{65}$	C1
A2	$\sqrt{25}=5$	$\sqrt{18}=4.24$	$\sqrt{10}$	C3
A3	$\sqrt{36}=6$	$\sqrt{25}=5$	$\sqrt{53}=7.28$	C2
A4	$\sqrt{13}=3.61$	0	$\sqrt{52}=7.21$	C2
A5	$\sqrt{50}=7.07$	$\sqrt{13}=3.61$	$\sqrt{45}=6.71$	C2
A6	$\sqrt{52}=7.21$	$\sqrt{17}=4.12$	$\sqrt{29}=5.38$	C2
A7	$\sqrt{65}=8.06$	$\sqrt{52}=7.21$	0	C3
A8	$\sqrt{5}=2.24$	$\sqrt{2}=1.41$	$\sqrt{58}=7.62$	C2

After 1 epoch, new clusters become.

C1 : A1

C2 : A3, A4, A5, A6, A8.

C3 : A2, A7.

New cluster centres:

C2: $x: \frac{8+5+7+6+4}{5} = \frac{30}{5} = 6.$

$y: \frac{4+8+5+4+9}{5} = \frac{30}{5} = 6.$

New centre = (6, 6)

$C3 = \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = (1.5, 3.5)$

Ans(i)

After 1 epoch, centres are

C1 : (2, 10)

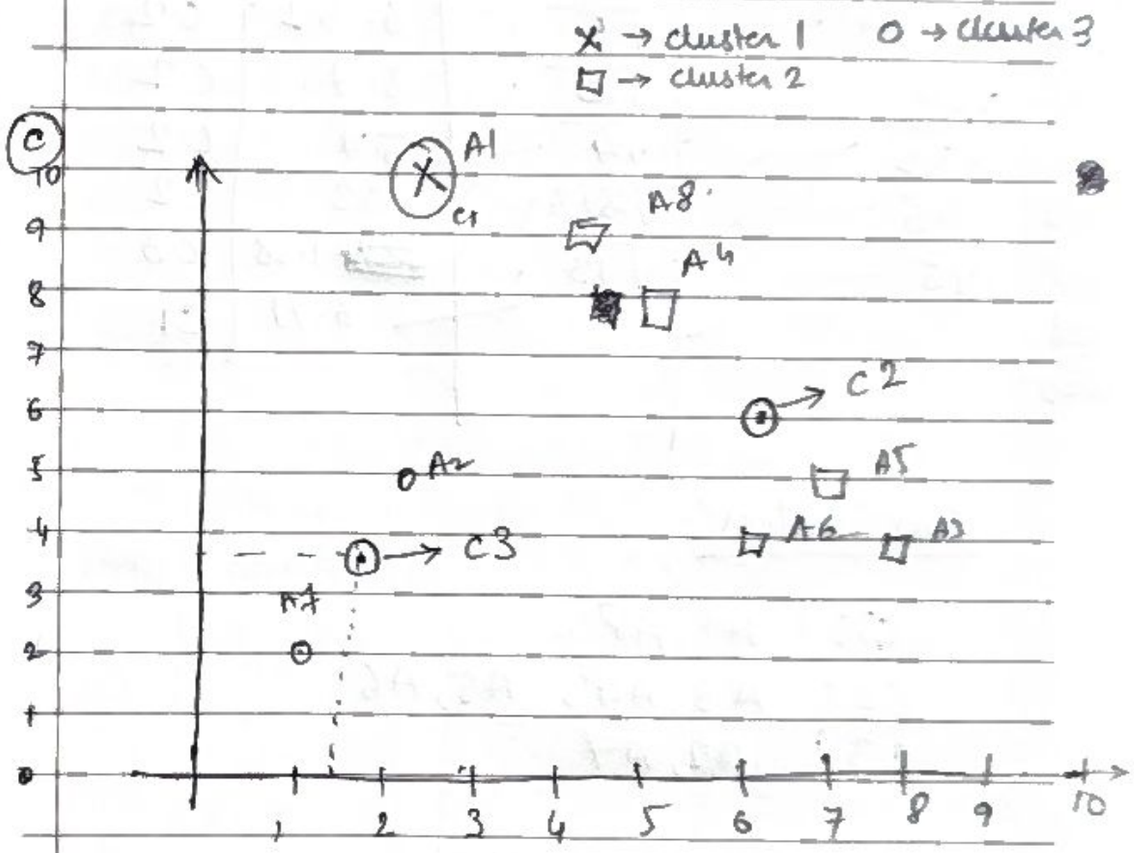
C2 : (6, 6)

C3 : (1.5, 3.5)

and clusters are

(A1), (A3, A4, A5, A6, A8)

and (A2, A7).



C1, C2, C3 are the cluster centres.

DATE

PLACE

Epoch 2

Pt	Dist to C1	Dist to C2	Dist to C3	Assigned cluster
A1	0			
A2	$\sqrt{25}=5$	$\sqrt{16+16}$	$\sqrt{42.25}$	C1
A3	$\sqrt{72}=8.49$	$\sqrt{8}$	1.58	C3
A4	$\sqrt{13}=3.61$	$\sqrt{5}$	6.52	C2
A5	$\sqrt{50}$	$\sqrt{2}$	5.7	C2
A6	$\sqrt{52}$	$\sqrt{4}$	5.7	C2
A7	$\sqrt{145}$	$\sqrt{25+16}$	4.53	C2
A8	$\sqrt{5}$	$\sqrt{13}$	1.58	C3
			6.71	C1

New Clusters:

C1: A1, A8

C2: A3, A4, A5, A6.

C3: A2, A7.

New centre means:-

C1 $\Rightarrow$ $\frac{2+4}{2}, \frac{10+9}{2} = (3, 9.5)$

C2 $\Rightarrow$ $\frac{8+5+7+6}{4}, \frac{4+8+5+4}{4} = (6.5, 5.25)$

C3 $\Rightarrow$ $(1.5, 3.5)$.

DATE

PLACE

Compute distances again.

Pt	C1	C2	C3	Cluster
A1	1.80	6.36	6.5	C1
A2	4.6	4.76	1.58	C3
A3	6.72	2.06	~6.52	C2
A4	2.5	3.04	~5.7	C1 ← Changed
A5	4.84	1.45	~5.7	C2
A6	6.02	1.53	~4.53	C2
A7	8.83	6.2	1.58	C3
A8	1.12	4.2	67	C1

New Clusters! C1 → A1, A4, A8.

C2: A3, A5, A6.

C3: A2, A7,

New centres:

C1 = $\frac{2+5+4}{3}, \frac{10+8+9}{3} = (3.67, 9)$

C2 = $\left(\frac{8+7+6}{3}\right), \frac{4+5+4}{3} = (7, 4.33)$

C3 = $(1.5, 3.5)$ Same

Thus, convergence is reached after epoch 4.

June

DATE

PLACE

Final cluster Results

Epoch	cluster 1	cluster 2	Cluster 3
1	A1	~~A3, A4, A5, A6, A8~~	~~A2, A7~~
2	A1, A8	A3, A4, A5, A6	A2, A7
3	A1, A4, A8	A3, A5, A6	A2, A7
4	A1, A4, A8	A3, A5, A6	A2, A7

DATE

PLACE

(e) We need to prove

$$\sum_{i=1}^{m} \|z_i - z\|^2 \geq \sum_{i=1}^{m} \|z_i - \bar{z}\|^2$$

This is equivalent to saying mean $\bar{z}$ minimizes the sum of squared distances from points z_i. Any other point gives greater or equal distance squared.

we know,

$$\text{mean } \bar{z} = \sum_{i=1}^{m} z_i$$

z = arbitary number.

using an identity,

$$\|a-b\|^2 = \|a-c+c-b\|^2$$

$$= \|a-c\|^2 + 2\langle a-c, c-b\rangle + \|c-b\|^2$$

June

When, $a = z_i - \bar{z}$

$b = \bar{z}$

$c = \bar{z}$.

$$\sum_{i=1}^{m} \|z_i - z\|^2 = \sum_{i=1}^{m} \|z_i - \bar{z}\|^2 + 2\sum_{i=1}^{m} \langle z_i - \bar{z}, \bar{z} - z \rangle + \sum_{i=1}^{m} \|\bar{z} - z\|^2 .$$

Now,

$$\sum_{i=1}^{m} (z_i - \bar{z}, \bar{z} - z)$$

$$= \sum_{i=1}^{m} \langle (z_i - \bar{z}), (\bar{z} - z) \rangle$$

$$= \left\langle \sum_{i=1}^{m} (z_i - \bar{z}), \bar{z} - z \right\rangle$$

$$= \langle 0, \bar{z} - z \rangle = 0 .$$

Mean balances out the data pts. So sum of all offsets is always 0.

$$\sum_{i=1}^{m} \|\bar{z} - z\|^2 = m \cdot \|\bar{z} - z\|^2$$

Thus, expression boils down to,

$$\sum_{i=1}^{m} \|z_i - z\|^2 = \sum_{i=1}^{m} \|z_i - \bar{z}\|^2 + m \cdot \|\bar{z} - z\|^2$$

$$\geq \sum_{i=1}^{m} \|z_i - \bar{z}\|^2 .$$

with equality iff when $z = \bar{z}$

The second term is always non-negative since, its squared norm:

(f) The process is a description of K-means with a greedy initialization.

1st centre: random
2nd centre: pick point furthest from existing centers.
After that standard K-means.

At each step,

Given centroids u^t, each pt is assigned to nearest centroid.
Here, sum of square errors doesn't increase as each pt moves closer to its actual centre.

In update step,

Given C^{t+1} clusters, we recompute new centroids u^{t+1} as the mean of each cluster.

Earlier we proved, mean minimises sum of squared distances.

Thus, SSE decreases or remains same

$$SSE(C^{t+1}, u^{t+1}) \leq SSE(C^t, u^t)$$

Thus, as long as the clustering changes, we will move towards convergence, with greedy initialization.

Greedy initialization tries to spread out centres in the start preventing any kind of skewed choice.

This results in a lower iterations need to converge, since we have reduced chances of local bad minima when compared to random initialization

(c) Each point can belong to any one of the K- clusters.
Thus, K^n possible ways to assign n points to K clusters.
Thus, each iteration has either to decrease SSE strictly (until convergence).

DATE

PLACE

Since, there are finitely, many clusterings, it has to terminate in finite loop. Thus, gurrantee termination.

DATE 11/04/25
PLACE

(4) Given, $K = 2$ for a dataset with 4 points

(a) & GMM to be fit.

at t^{th} time step, we have

θ^t with following parameters:

$\pi_1 = 0.3 \quad \pi_2 = 0.7$.

$\mu_1 = 2 \quad \mu_2 = 3$

$\sigma_1^2 = 1 \quad \sigma_2^2 = 1$.

$f(x_i | z_i = 1)$ & $f(x_i | z_i = 2)$ for $i = 1$ to 4 are given.

In E step, we are calculating responsibilities λ_i^K, the probability that point x_i belongs to cluster K.

z_i → represents the latent variable that indicates which ~~cluster~~ Gaussian component generated the data point x_i

DATE
PLACE

Responsibility of cluster K for point i is

$$\lambda_i^K = \frac{\pi_k \cdot f(x_i | z_i = K)}{\sum_{j=1}^{K} \pi_j \cdot f(x_i | z_i = j)}$$

Responsibility of cluster 2 from point 1 ie. how likely is that point 1 comes from cluster 2 is

$i = 1, \quad K = 2$

$$\lambda_1^2 = \frac{(0.7)(0.054)}{0.3 \times 0.242 + 0.7 \times 0.054}$$

$$= \frac{0.0378}{0.0726 + 0.0378} = 0.3423$$

So, probability of point 1 belonging to cluster 1 is 34.23%.

DATE

PLACE

(b) If we pause after ~~cluster~~ E step, we have all the responsibilities λ_i^k.

If we do a hard assignment, we simply shall assign each point to the cluster for which it has highest responsibility.

ie $z_i = \arg\max_k \lambda_i^k$

Since, $\lambda_1^2 = 0.3423$

$\lambda_1^1 = 1 - 0.3423 = 0.6577$

Among this, max is cluster 1.

So, $\boxed{z_1 = 1}$

DATE

PLACE

Point 2

cluster 1 point 2

$$\lambda_2^1 = \frac{\pi_k f(x_i|z_i=k)}{\sum_{j=1}^{k} \pi_j \cdot f(x_i|z_i=j)}$$

$$= \frac{\pi_1 f(x_2|z_2=1)}{\pi_1 f(x_2|z_2=1) + \pi_2 f(x_2|z_2=2)}$$

$$= \frac{0.3 \times 0.399}{0.3 \times 0.399 + 0.7 \times 0.242}$$

$$= \frac{0.1197}{0.2891} = 0.4141$$

Here, $\lambda_2^2 = 1 - 0.4141$

$> \lambda_2^2$

thus $\boxed{z_2 = 2}$.

DATE

PLACE

$$\lambda_3^1 = \frac{\pi_1 f(x_3 | z_3 = 1)}{\pi_1 f(x_3|z_3=1) + \pi_2 f(x_3|z_3=2)}$$

$$\doteq \frac{0.3 \times 0.242}{0.3 \times 0.242 + 0.7 \times 0.399}$$

$$= \frac{0.0726}{0.3519} = 0.2063.$$

$$\lambda_3^2 = 1 - 0.2063 = 0.7937$$

Thus, ~~etc~~ $\boxed{z_3 = 2}$.

Point 4

$$\lambda_4^1 = \frac{\pi_1 f(x_4 | z_4 = 1)}{\pi_1 f(x_4|z_4=1) + \pi_2 f(x_4|z_4=2)}$$

$$= \frac{0.3 \times 0.054}{0.3 \times 0.054 + 0.7 \times 0.242}$$

$$= \frac{0.0162}{0.1856} = 0.0873$$

DATE

PLACE

$\lambda_4^2 = 1 - \lambda_4^1 \geq \lambda_4^1$.

Since $\lambda_4^1 = 0.0873$

thus $z_4 = 2$.

Final vector notation of cluster

$= [1 \ 2 \ 2 \ 2]$